MINIMALLY BIASED WEIGHT DETERMINATION IN PERSONNEL SELECTION

Alan Jessop

Durham Business School
Mill Hill Lane
Durham
DH1 3LB

e-mail: a.t.jessop@durham.ac.uk
phone: 0191 374 1233
fax: 0191 374 3748

MINIMALLY BIASED WEIGHT DETERMINATION IN PERSONNEL SELECTION

Abstract

The derivation of weights from preference statements is subject to difficulties, some of which are due to the unreliability of the judgement of the decision maker. To overcome this Jaynes' principle of maximum entropy has been invoked and may be applied either to weights or to the linear weighted scores of the candidates in a selection problem. When candidates are relatively few the two strategies give different styles of interaction. These are discussed and illustrated by application to a problem of personnel selection.

KEY WORDS: multiple criteria analysis, human resources, entropy, personnel selection

1. Introduction

The use of simple weighted sums as a method of aggregating incommensurate characteristics or alternatives is common (Stewart, 1992) and has generated a vast literature. The model is

$$y_i = \sum_j w_j u_j(x_{ij}) \tag{1}$$

where x_{ij} is the value of attribute j achieved by choosing alternative i; w_j is a weight; $u_j(\cdot)$is a value function and

$$\sum_j w_j = 1 \tag{2}$$

The vector of scores, Y, provides a basis for ranking or selection.

It is generally perceived that the largest technical difficulty is finding values for the weights and many methods are available for deriving weights from preference judgements. The choice of method may be more a matter availability or of personal preference on the part of the analyst than on any other consideration (Bottomley and Doyle, 2001). These judgements are necessarily subjective, and so fallible, as a result of the cognitive limits of the respondent (Barron and Barrett, 1996; Borcherding, Schmeer and Weber, 1995; Ranyard and Abdel-Nabi, 1993; Larichev, 1992; von Winterfeldt and Edwards, 1986). Inasmuch as these result in biases which, though psychological in origin and innocent of any prejudice, may be (mis)taken as evidence of behaviour which is socially unacceptable or even illegal then the derivation of weights which may be seen as unbiased becomes a pressing task. Nowhere is this more true than in personnel selection.

In what follows two articulations of unbiased weight estimates are given; issues surrounding personnel selection are reviewed; an illustrative application of unbiased estimation to a personnel problem is discussed.

2. Avoiding bias

Edwin Jaynes was concerned to provide estimates of probability distributions which were unbiased in that they were as uniform (flat) as possible subject only to specified constraints. He proposed that maximising the entropy of the distribution subject to these constraints provided just such an estimate, the maximisation ensuring the flattest possible distribution (Jaynes, 1957). These flat distributions are said to be minimally discriminating; in Jaynes' case minimally discriminating between the relative likelihoods of different events or states. The same principle may be applied to the calculation of weight distributions and a review is provided by Jessop (1999). This method is therefore proposed as a way of finding weights and so making selection decisions which may be said to be unbiased.

The entropy of an arbitrary vector, Z, is

$$H(Z) = \ln(\sum_i z_i) - \sum_i z_i \ln(z_i) / \sum_i z_i \tag{3}$$

which becomes

$$H(Z) = -\sum_j z_j \ln(z_j) \tag{4}$$

in those cases, such as probabilities and weights, in which the sum of the vector is unity.

The most common application of this method in multicriteria modelling is in the derivation of weights directly. Some preference information may be available, as ratios between weights or as the resolution of bicriterial problems, and these provide constraints given which, and with (2), $H(W)$ is maximised. The full programme is:

$$\max \quad H(W) = -\sum_j w_j \ln(w_j) \tag{5}$$

$$\text{s.t.} \quad \sum_j w_j = 1 \tag{6}$$

$$w_i \geq \varepsilon \; ; \quad \forall\, i \tag{7}$$

$$w_i \geq a w_k \tag{8}$$

$$w_i = b w_k \tag{9}$$

$$\sum_j w_j u_j(x_{ij}) \geq c \sum_j w_j u_j(x_{kj}) \tag{10}$$

$$\sum_j w_j u_j(x_{ij}) = d \sum_j w_j u_j(x_{kj}) \tag{11}$$

Constraint (7), which is optional, ensures that all weights are non-zero. In the calculations described below $\varepsilon = 0.02$. Constraints (8) and (9) encode the results of pairwise comparisons of weights and constraints (10) and (11) similarly encode the results of pairwise comparisons of candidates. Parameters *a* to *d* are given by the decision maker in those comparisons in expressing the strength of the preferences expressed. There are as many of the constraints (7) to (11) as the decision maker wishes to provide.

Call this method *maxEntWeights*. In the absence of preference information the result is a uniform weight distribution.

Calculating weights in this way will in general give scores, *Y*, different for each candidate and so will permit a decision to be made between them. But, as has been pointed out (Jessop, 1999), it seems counterintuitive to say that minimally discriminating weights yet permit discrimination. The conflict arises as the result of a tension between the dual foci of weights and scores. If the purpose of the task is to find weights which reflect the underlying values of the decision maker then using *maxEntWeights* may be justifiable. However, in most cases, and certainly in personnel selection, this is not the purpose: what is required is the least biased selection from among candidates. That being so it makes sense to be minimally discriminating between candidates rather than between weights and so to maximise $H(Y)$. The programme is as given above but with $H(Y)$ replacing $H(W)$. The objective function is, from (1) and (3),

$$\max H(Y) = \ln\Big(\sum_i\sum_j w_j u_j(x_{ij})\Big) - \sum_i\Big[\sum_j w_j u_j(x_{ij}) \ln\Big(\sum_j w_j u_j(x_{ij})\Big)\Big] \Big/ \sum_i\sum_j w_j u_j(x_{ij}) \qquad (12)$$

with the constraints (6) to (11) as before. Call this method *maxEntScores*.

It is unlikely, and undesirable, that in practice a decision maker would proceed without preference but, whatever preference information is given, it will remain the case that *maxEntScores* will give a result that is minimally discriminating between candidates and *maxEntWeights* will not and that in personnel selection this is a desirable precaution against the effects of judgemental bias.

The results obtained by using *maxEntScores* will depend in an interesting way on both the number of attributes and the number of candidates. When the number of attributes is large compared to the number of candidates it will be possible to incorporate some judgmental constraints without being able to discriminate between alternatives. When the number of candidates is greater than the number of attributes then, even with no judgmental information, a preference order will be induced among candidates. The effect is basically due to different degrees of freedom, though the exact point corresponding to the zero degrees of freedom situation may be difficult to determine in advance if the judgmental constraints are expressed as inequalities rather than equalities. Call these two types of problem *Short Lists* and *Long Lists*, in recognition of the two stages of most selection problems, not least in personnel selection.

3. Personnel selection

The assessment and selection of personnel for promotion, reassignment or to fill new positions is a task much discussed and is the subject of a large body of research. A good review is given by (Robertson and Smith, 2001) who describe the commonly used process:

1. detailed analysis of the job leading to
2. an indication of the psychological attributes required of a successful candidate
3. personnel selection methods aim to assess the extent to which these attributes are possessed by candidates
4. a validation process tracks the success of the selection process in identifying suitable candidates.

Those factors assessed will typically be both job specific and generic in that they refer to the personality of the candidate, though Ree, Earles and Teachout (1994) suggest that job specific evaluations offer little that could not otherwise be found by an assessment of general intelligence. Personality is generally described by the "Big Five" personality factors of emotional stability, extraversion, openness, agreeableness and conscientiousness (Salgado, 1997).

Bratton and Gold (1994) say that the objectives of a selection process are twofold: to assess the differences between candidates and to predict future performance. The extent to which a process is successful may be measured by its reliability (the extent to which results may be replicated by, for instance, different assessors) and validity (the degree to which it measures what it says it will). This latter is difficult to assess due to the large samples required and the various other temporal changes that may affect employees. Table 1 shows the validity of some criteria. Table 2 shows an assessment of the reliability of some methods, assessed as the improvement that they offer over random selection of candidates. The use of different methods in some European countries is shown in Table 3. Similar results were given by Robertson and Makin (1986).

The difference between the popularity of methods used and their evidential value is striking and is the cause for dissatisfaction of writers in this area (Guion, 1998b). Two useful trends should be noted. First, the use of structured interviews rather than the more freewheeling conversations which once were the more common (and still may be) has increased the validity of the interview (Robertson and Smith, 2001; Cortina et al, 2000). However, the extent to which interviews provide incremental information given that other assessments, notably of cognitive ability and conscientiousness, are available is disputed: Cortina et al (2000) present evidence that highly structured interviews can contribute "substantially to prediction" of performance in the job, while Barrick, Patton and Haugland (2000) think that it does not, although it may be well suited to assess the fit of a candidate within an organisation. Second, several techniques are used in assessment centres and assessments made which typically estimate performance against more criteria than the there are tests. Henderson, Anderson and Rick (1995) describe a centre of four exercises from which performance against fourteen criteria was assessed. The centre described by Blackham and Smith (1989) used six tests or exercises to evaluate ten attributes. In both cases ratings were given on a simple scale (five point in the latter) and aggregated using equal weights. Lowry (1994) gives evidence that structured interviews and assessment centres have similar validity.

4. Formal methods

The use of scoring or rating to describe assessments of performance against criteria is not limited to assessment centres. Scoring in structured interviews is also common: 63% of the respondents in a study by Barclay (2001) used scoring for this purpose. The motivation for explicit scoring systems, as for any structured method, is to reduce the effects of interviewer subjectivity and bias (Cascio, 1991, ch 5; Campion, Palmer and Campion,1997; Guion, 1998a, ch 12). While the articulation of assessment judgement via scores is accepted the use of a similarly explicit aggregation of a number of scores to give an overall judgement of the merit of candidates seems to be resisted and so the benefits obtained from scoring are not fully utilised (Barclay, 2001). This may in part be due to the perceptions of interviewers that interviews, however strongly structured, do not have the scientific base of, say, psychometric tests and so are not capable of giving numerically codable results. Alternatively, it may be that interviewers and managers wish to maintain the flexibility in weighting the importance of different components that comes with inexplicit combination. The use of an aggregating model more complicated than just summing the individual ratings, as in the case of the assessment centres discussed above, will not be helped by the generally low use of IT by personnel specialists (Huo and Kearns, 1992; Kinnie and Arthurs, 1996).

Nonetheless, there have been some attempts to apply formal methods to the aggregation problem. Fuzzy set theory has been proposed (Miller and Feinzig, 1993; Karsak, 2001; Capaldo and Zollo, 2001) as have the applications of linear weighted sums familiar in multicriteria analyses (Bohanec, Urh and Rajkovič,1992; Gardiner and Armstrong-Wright, 2000; Spyridakos et al, 2001). Timmermans and Vlek (1992; 1996) explore how the use of formal methods may help in overcoming bias and information overload and conclude that using a weighted sum for aggregation is of use for some problems but that it is unnecessary for small problems and fails to be of help in large complex problems. Roth and Bobko (1997) review some of the issues surrounding the use of multiattribute methods in human resource management. Ganzach, Kluger and Klayman (2000) give a method of combination based on regression analysis.

This last study compares a single unaided assessment of candidates and what is effectively a weighted sum and finds no great difference between the two, although the authors note that this finding is somewhat atypical and cite other studies showing the superiority of more formal disaggregated methods. But increased accuracy is not the only objective. The elimination of bias is becoming increasingly important not only in the technical sense of reducing the impact of cognitive bias but in eliminating, and being seen to eliminate, gender bias, racial bias and other similar biases which are socially undesirable or illegal. Gardiner and Armstrong-Wright (2000) show how a multicriteria approach to such decisions can be of help by providing a procedure which is fair, consistently applied, well documented, transparent and so legally defensible. They also note that personnel decisions are not often made by one individual and that the problem is essentially one of group decision support. In these potentially litigious situations it seems reasonable that minimising discrimination between candidates by using *maxEntScores* is most likely to guard against accusations of bias in favour of or against a particular candidate.

It is this property of being minimally discriminating between candidates except through explicitly stated judgements which makes *maxEntScores* such an appropriate decision aid for personnel selection, for avoiding implicit, even unconscious, bias is exactly what managers faced with these decisions are required to do. It is perhaps more common, if a little mistaken, to articulate this requirement by having equal or near equal weights: this is *maxEntWeights*.

5. Illustration

The main issues to arise from this discussion are, first, the usefulness of a weighted average scoring model in general and, second, the relative usefulness of *maxEntWeights* and *maxEntScores* as criteria in the determination of weights. While *maxEntScores* is the preferable criterion for avoiding bias in selection problems, the practical issue is whether the different style of interaction when using this criterion would be practicable or whether users would find the fact that discrimination between candidates did not immediately result from their first few judgemental inputs sufficiently counterintuitive to render the method unacceptable. To examine these issues a simple experiment was conducted.

The experimental group was sixteen staff and MBA students of Durham Business School. The staff were academics and administrators and the students were people who had held managerial positions before joining the MBA programme. Each was asked to select which of five candidates, a – e, should be chosen for the post of Postgraduate Admissions Officer at the University, a post which had in fact recently been advertised and filled. They were provided with the two page description of the job that had been sent to applicants. From the required skills described a list of seven criteria was constructed:

1. Written communication
2. Oral communication
3. Planning
4. Organising ability
5. Team player
6. Works independently
7. Decisiveness

The job description contained no indication of the relative importance of the criteria.

Additionally each person was given the one page description of the experiment shown in Figure 1 sufficiently in advance of the experiment itself that they could understand clearly what was to happen and what were the characteristics of the job.

The performance of each candidate against the criteria were described on a five point scale, with 5 denoting excellence: these are the values $u_j(x_{ij})$ in (1). These values were generated randomly, with adjustment to ensure that a large number of very poor scores of 1 was avoided because the presence of too many such weak performances might lead to an early elimination of candidates without the need to make the tradeoff judgements required by the methods, *maxEntWeights* and *maxEntScores*, being tested. Two sets of hypothetical candidates, called the Red and Blue sets, were used and are shown in Table 4. As can be seen from the sums given in the last line of the table, candidates in the Red set were slightly the stronger and exhibited a little higher spread of abilities between candidates.

The size of the problem is seen as within the range discussed by Timmermans and Vlek (1992; 1996) within which multiattribute decision aids ma be of use.

The experiment was to test the two forms of weight calculation and the corresponding interactions. To do this each of the participants in the experiment was asked to make two selections, one using *maxEntWeights* and one using *maxEntScores*. Since there was likely to be some, perhaps some considerable, learning carried over from the first selection task to the second, the order in which the two methods were used was varied. Each selection also used a different candidate data set. As a result there were four task pairs, each of which is coded according to the first task, as shown in Table 5.

Each participant was confronted with a computer screen showing the display as described in Figure 1 but with weighted average scores at the foot of each column. Initially all weights were set to zero and so these scores were also zero. It was explained that this was so because no initial judgemental information had yet been given and that scores would appear as soon as the first pairwise statement had been made. After each pairwise statement the corresponding constraint was entered and the weights recomputed and calculated scores displayed. This process continued until a recommendation could comfortably be made. The time taken for the task was noted. Each participant gave a score indicating how difficult they found the task with 1 representing "easy" and 5 "hard". Comments both about the tasks and about the approach in general were also optionally provided.

The task was to select rather than to find weights *per se*. It has been noted elsewhere, by Jessop (2002) for instance, that people nonetheless like to focus on weights as these are seen as important in terms of policy or as explicit representations of personal preferences.

It was thought to be of some interest to examine this effect here and so to this end values of weights were at no point shown.

No restrictions were placed on how each person should interact with the data so that statements about criteria or scores could be given in any order and either singly or in groups. It was also permissible to eliminate candidates from further consideration at some stage in the process if, for instance, it was believed that performance against a crucial criterion so poor as to disbar the candidate from the job.

In the instructions (Figure 1) two typical responses were given to indicate that preference statements about criteria or candidates or both were permissible. Although the example given for criteria was phrased to elicit an equality constraint most responses were inequality statements such as "written communication is more important than planning" rather than the specification of a precise ratio between the weights. This was so for preference statements concerning both criteria and candidates.

6. Results

The results are summarised in Table 6. Column *a* gives the identifying number of the respondent; column *b* gives the task pairs; columns *c* to *f* give the times and difficulty scores for the first and second problems (P1 and P2 respectively). Columns *g* and *h* show the judgemental inputs provided, with those shown in parentheses for P2 indicating judgements carried over from P1 so that, for instance, in P1 respondent 6 (hereafter written as {6}) provided 2 judgemental inputs regarding weights directly (e.g. "I think criterion 4 is at least twice as important as criterion 2") and 1 regarding candidates ("I believe that candidate *c* is better than candidate *e*") all three of which were used as constraints in the optimisation. In P2 the two judgements about criteria were carried over from P1 and an additional judgement made about criteria and one about candidates. It is, of course, impossible to carry over a weight about candidates since the set of candidates changes. In all cases except one {16} all judgemental constraints were carried over from P1 to P2.

Table 7 shows the mean characteristics for each pair group. With only four responses in each group a detailed statistical analysis is inappropriate but some conclusions may, with circumspection, be drawn.

The number of extra judgements needed for P2 is small in all cases and will not be further considered here.

The four experimental pairs have been ordered according to the number of judgements needed for P1. The two *maxEntScores* problems head the list. This is as expected since this model requires more constraints before non-uniform scores are found. In both *maxEntScores* and *maxEntWeights* problems the Blue set required less input than the Red for a satisfactory recommendation. As seen at the bottom of Table 4 the Blue set are generally poorer performers and also less differentiated so, again, the results are as expected.

Both time taken and the perceived difficulty of the P1 problems are roughly negatively correlated when considering means though the picture is less clear when all sixteen data are considered as, for instance, in Figure 2. The same picture emerges, though less strongly, for P2 problems. Inasmuch the idea of a negative correlation is entertained, however tentatively, it suggests that the extra number of judgements evidently required by the method of weight calculation, moderated by the inherent characteristics of the data, have the benefit of requiring less deliberation by the decision maker on other grounds.

In Table 7 the P2 problems are the complements of the P1 problems (Table 5). In all four groups the mean times fall substantially from P1 to P2. The data are represented in Table 8. Here the P2 problems are not the complements of P1 but are the same method/data pair in both cases. For example, BS problems are P1 in the group labelled "BS" and P2 in the group labelled, complementarily, "RW". In Table 8 both have been entered on the "BS" line. The groups which require most and least input, BS and RW, may be thought to be the hardest and simplest problems respectively as judged by the information required for a decision to be made. Looking at these two in Table 8 it is seen that for the hard problem, BS, the solution time is not much different whether it was P1 or followed, as P2, the simplest problem, RW. In this second case, because of the simplicity of the first of the pair not much was learned that was useful in tackling the harder problem which required greater informational input. Look now at the RW entries and exactly the opposite is seen: when RW is encountered as P2 after BW then the mean time is just 3.25 mins., a large reduction from 11.75mins. RS and BW problems have similar information requirements and so this effect is not seen when comparing their times. The same observations hold when considering the difficulty scores but, since the measure here is a coarse five point scale, the magnitude of the effects is smaller.

In all sixteen cases judgements about criteria preceded those about candidates and were more numerous. Most participants began by comparing the relative importance of the two most important criteria. Some then continued comparing the second and third most important and so on. In large part these constraints were inequalities establishing a rank ordering of criteria and candidates.

Although weights were not displayed nobody asked what values the weights had, being quite content to just input constraints. In just two cases when using *maxEntScores* respondents remarked that the scores were unchanged after some information had been given. It was explained that this was because not enough information had been provided to permit distinctions to be made and this was accepted, without further question, as a satisfactory explanation: it was also true.

Almost all participants liked the method for its objectivity and as a way both of removing potential bias and as a means for ensuring that all performance data were examined. Most made the point that the general method was good for filtering but that a final decision would have to be based on, or confirmed by, a knowledge of the applicants and spoke warmly of the role of the interview particularly in determining whether a candidate would fit well into the organisation. One participant {8} put it this way: "the scores measure who can do the job best but an interview finds who is best for the job". Being told of the poor properties of unstructured interviews did not cause revisions of these views.

Three participants {2}, {3} and {15} felt that performance assessments of 1 were unacceptable and so ruled out those two candidates with these low assessments, thereby reducing the selection problem to a choice between three candidates. Others noted these low values but said that they would assume that appropriate staff development would improve performance to an acceptable level and so selected a candidate with an assessment of 1. The majority of participants made no comment about this issue. One participant {7} would have liked to see an unweighted sum for each candidate, as given in Table 4, as a guide to choice.

7. Conclusion

Two articulations of unbiasdness in the context of multiattribute decision making have been described and applied to an illustrative problem of personnel selection. Both were shown to be feasible. The preferred method, *maxEntScores*, is shown to be practicable and that the requirement of some small number of judgmental inputs before it is possible to begin to differentiate candidates caused no difficulty for users. This initial response lag is the main difference in the styles of interaction required by the two methods. While this start up cost, as it might be seen, may at first seem unusual it ought not to be surprising that some extra effort is necessary in final selection in order that the process may be seen as free from bias. In any case, when making selections in practice it would never be sufficient to provide just one judgmental input and then be able to feel comfortable with the result. In this respect the initial inertia or stickiness of *maxEntScores* is much closer to the experience of decision makers than the alternative concentration on weights and the instant differentiation it allows. This characteristic of forcing more initial inputs appears to reduce overall solution time and perceived difficulty.

The experiment was artificial: a real selection process was not observed. The participants did not have to live with the consequences of their decisions in the way that decision makers often do. Selecting participants familiar with interviewing or the University or both and using a real job description (though not, for reasons of confidentiality data on real applicants) added a degree of realism. Nonetheless, it was a constructed experiment and so the results should be interpreted circumspectly. Some results, incuriosity about the values of weights, for instance, should be viewed in this light.

The experimental problem represents an extreme case in that the selections were made in ignorance of candidates and without participation in the assessment of performance against individual criteria. This certainly would prevent accusations of bias but is an unlikely, and probably an undesirable, procedure: those making the final recommendations will usually have been part of all stages of the selection process. It is clear from the respondents that having some appreciation of the person behind the figures is important, as probably it ought to be, but the unjustified faith which some had in the utility of interviews suggests that the need for a decision aid which guards against unjustified bias is as great as ever.

References

Barclay, J.M., 2001. Improving selection interviews with structure: organisations' use of "behavioural" interviews, Personnel Review 30 81–101.
Barrick, M.R., Patton, G.K., Haugland, S.N., 2000. Accuracy of interviewer judgements of job applicant personality traits, Personnel Psychology 53 925–951.
Barron, F.H., Barrett, B.E., 1996. The efficacy of smarter – Simple Multi-Attribute Rating Technique Extended to Ranking, Acta Psychologica 93 23–36.
Beardwell, I., Holden, L., 1994. Human Resource Management: A Contemporary Perspective, Pitman, London.
Blackham, R.B., Smith, D., 1989. Decision-making in a management assessment centre, Journal of the Operational Research Society 40 953–960.
Bohanec, M., Urh, B., Rajkovič, V., 1992. Evaluating options by combined qualitative and quantitative methods, Acta Psychologica 80 67–89.
Borcherding, K., Schmeer, S., Weber, M., 1995. Biases in multiattribute weight elicitation. In: Caverni J-P., Bar-Hillel M., Barron FH., Junngermann H. (Eds), Contributions to Decision Making – I, Elsevier, Amsterdam.
Bottomley, P.A., Doyle, J.R., 2001. A comparison of three weight elicitation methods: good, better and best, Omega 29 553–560.
Bratton, J., Gold, J., 1994. Human Resource Management: Theory and Practice, Macmillan, Basingstoke.
Campion, M.A., Palmer, D.K., Campion, J.E., 1997. A review of structure in the selection interview, Personnel Psychology 50 655–702.
Capaldo, G., Zollo, G., 2001. Applying fuzzy logic to personnel assessment: a case study, Omega 29 585–597.
Cascio, W.F., 1991. Applied Psychology in Personnel Management (3rd edn), Prentice-Hall, Englewood Cliffs NJ.
Cortina, J.M., Goldstein, N.B., Payne, S.C., Davison, H.K., Gilliland, S.W., 2000. The incremental validity of interview scores over and above cognitive ability and conscientiousness scores, Personnel Psychology 53 325–351.
Dany. F., Torchy, V., 1994. Recruitment and selection in Europe: policies practices and methods. In: Brewster C., Hegewisch A. (Eds), Policy and Practice in European Human Resource Management, Routledge, London.
Ganzach, Y., Kluger, A.N., Klayman, N., 2000. Making decisions from an interview: expert measurement and mechanical combination, Personnel Psychology 53 1–20.

Gardiner, A.R., Armstrong-Wright, D., 2000. Employee selection and anti-discrimination law: implications for multi-criteria group decision support, Journal of Multi-Criteria Decision Analysis 9 99–109.
Guion, R.M., 1998a. Assessment, Measurement and Prediction for Personnel Decision, Lawrence Erlbaum, Mahwah NJ.
Guion, R.M., 1998b. Some virtues of dissatisfaction in the science and practice of personnel selection, Human Resource Management Review 8 351–365.
Henderson, F., Anderson, N., Rick, S., 1995. Future competency profiling, Personnel Review 24 19–31.
Huo, Y.P., Kearns, J., 1992. Optimizing the Job-person Match with Computerized Human Resource Information Systems, Personnel Review 21 3–18.
Jaynes, E.T., 1957. Information theory and statistical mechanics, Phys. Review 106 620–630 and 108 171–190.
Jessop, A., 1999. Entropy in multiattribute problems, Journal of Multi-Criteria Decision Analysis 8 61–70.
Jessop, A., 2002. Prioritisation of an IT budget within a local authority, Journal of the Operational Research Society 53 36–46.
Karsak, E.E., 2001. Personnel selection using a fuzzy MCDM approach based on ideal and anti-ideal solutions. In: Köksalan M., Zionts S., (Eds), Multiple Criteria Decision Making in the New Millenium, Springer, Berlin.
Kinnie, N.J., Arthurs, A.J., 1996. Personnel specialists' advanced use of information technology: evidence and explanations, Personnel Review 25 3–19.
Larichev, O.I., 1992. Cognitive validity in design of decision-aiding techniques, Journal of Multi-Criteria Decision Analysis 1 127–138.
Lowry, P.E., 1994. The structured interview: an alternative to the assessment center, Public Personnel Management 23 201–215.
Miller, G.M., Feinzig, S.L., 1993. Fuzzy sets and personnel selection: discussion and an application, Journal of Occupational and Organizational Psychology 66 163–169.
Ranyard, R., Abdel-Nabi, D., 1993. Mental accounting and the process of multiattribute choice, Acta Psychologica 84 161–177.
Ree, M.J., Earles, J.A., Teachout, M.S., 1994. Predicting job performance: not much more than *g*, Journal of Applied Psychology 79 518–524. .
Robertson, I.T., Makin, P.J., 1986 .Management selection in Britain: a survey and critique, Journal of Occupational Psychology 59 45–47.
Robertson, I.T., Smith, M., 2001. Personnel selection, Journal of Occupational and Organizational Psychology 74 441–472

Roth, P.L., Babko, P., 1997. A research agenda for multi-attribute utility analysis in human resource management, Human Resource Management Review 7 341–369.
Salgado, J.F., 1997. The five factor model of personality and job performance in the European Community, Journal of Applied Psychology 82 30–43.
Spyridakos, A., Siskos, Y., Yannacopoulos, D., Skouris, A., 2001. Multicriteria job evaluation for large organisations, Eurpoean Journal of Operational Research. 130 375–387.
Stewart, T.J., 1992. A critical survey on the status of multiple criteria decision making theory and practice, Omega 20 569–586.
Timmermans, D., Vlek, C., 1992. Multi-attribute decision support and complexity: an evaluation and process analysis of aided versus unaided decision making, Acta Psychologica 80 49–65.
Timmermans, D., Vlek, C., 1996. Effects on decision quality of supporting multi-attribute evaluation in groups, Organizational Behavior and Human Decision Processes 68 158–170.
von Winterfeldt, D., Edwards, W., 1986. Decision Analysis and Behavioral Research, Cambridge University Press, Cambridge.

criterion	validity
cognitive ability and integrity	0.65
cognitive ability and structured interviews	0.63
cognitive ability and worksample	0.60
work sample tests	0.54
cognitive tests	0.51
structured interviews	0.51
job knowledge tests	0.48
integrity tests	0.41
personality tests	0.40
assessment centres	0.37
biodata	0.35
conscientiousness	0.31
references	0.26
years of job experience	0.18
years of education	0.10
interests	0.10
graphology	0.02
age	-0.01

Table 1. The validity of job performance criteria
(Robertson and Smith, 2001, Figure 1)
Note: validity has a maximum of 1.0

Method	*% better than chance*
Assessment centres	17–18
Work sample / simulation	14–29
Supervisory / line management evaluation	18
Mental ability	6–20
Biodata	6–14
References	3–7
Interviews	2–5
Personality test and self-report questionnaires	2.5
Graghology	0

Table 2. Selection method improvement on chance
(Nelson and Wedderburn in Beardwell and Holden, 1994, p.251)

	DW	*DK*	*E*	*F*	*FIN*	*IRL*	*N*	*NL*	*P*	*S*	*T*	*UK*
Application form	96	48	87	95	82	91	59	94	83	*na*	95	97
Interview panel	86	99	85	92	99	87	78	69	97	69	64	71
Bio data	20	92	12	26	48	7	56	20	62	69	39	8
Psychometric testing	6	38	60	22	74	28	11	31	58	24	8	46
Graphology	8	2	8	57	2	1	0	2	2	0	0	1
References	66	79	54	73	63	91	92	47	55	96	69	92
Aptitude test	8	17	72	28	42	41	19	53	17	14	33	45
Assessment centre	13	4	18	9	16	7	5	27	2	5	4	18
Group selection methods	4	8	22	10	8	8	1	2	18	3	23	13
Other	3	2	4	3	2	6	5	6	0	5	6	4

Table 3. Recruitment methods in 12 countries in Europe.
(% of selections using each) (Dany and Torchy, 1994)

	candidates (Red set)					candidates (Blue set)				
criterion	a	b	c	d	e	a	b	c	d	e
1	3	2	2	3	4	3	5	3	2	4
2	5	3	3	4	1	2	3	5	2	3
3	2	3	4	2	5	4	4	2	3	1
4	5	3	4	4	4	3	2	2	3	4
5	2	3	4	4	3	3	2	3	2	3
6	5	1	3	5	5	3	2	1	4	2
7	2	4	2	3	4	2	4	5	2	5
sum:	24	19	22	25	26	20	22	21	18	22

Table 4. Data sets used in experiment.

code	first problem, P1	second problem, P2
BS	Blue set & *maxEntScores*	Red set & *maxEntWeights*
RS	Red set & *maxEntScores*	Blue set & *maxEntWeights*
BW	Blue set & *maxEntWeights*	Red set & *maxEntScores*
RW	Red set & *maxEntWeights*	Blue set & *maxEntScores*

Table 5. The four experimental pairs

		time (mins)		difficulty		inputs	
no.	pairs	P1	P2	P1	P2	P1	P2
a	*b*	*c*	*d*	*e*	*f*	*g*	*h*
1	BS	11	1	2	1	W8	(W8)
2		5	4	1	2	W7,S1	(W7),S1
3		6	4	4	3	W12	(W12)
4		5	4	3	4	W6	(W6),S1
	mean:	6.75	3.25	2.5	2.5	W8.25,S0.25	(W8.25),S0.25
5	BW	10	7	2	3	W5,S1	(W5),S1
6		8	4	4	3	W2,S1	(W2),W1,S1
7		12	3	3	2	W4,S2	(W4),S1
8		20	2	5	3	W11	(W11)
	mean:	12.5	4	3.5	2.75	W5.5,S0.75	(W5.5),W0.25,S0.75
9	RS	9	5	2	3	W5	(W5),S1
10		12	5	2	4	W6	(W6)
11		6	3	2	4	W6	(W6)
12		5	1	2	1	W13	(W12)
	mean:	8	3.5	2	3	W7.5	(W7.5),S0.25
13	RW	19	8	3	4	W3,S2	(W3),W1,S2
14		8	1	4	1	W5	(W5)
15		12	9	4	4	W4	(W4),W4
16		8	7	4	5	W1,S1	S4
	mean:	11.75	6.25	3.75	3.5	W4.25,S0.5	(W4),W0.5,S0.5

Table 6. Results.

pairs	number of judgemental inputs for P1	number of extra inputs for P2	time		difficulty	
			P1	P2	P1	P2
BS	8.5	0.5	6.75	3.25	2.5	2.5
RS	7.5	0.25	8	3.5	2	3
BW	7.25	1	12.5	4	3.5	2.75
RW	4.75	1	11.75	6.25	3.75	3.5

Table 7. Average results for different pairs

pairs	number of judgemental inputs for P1	time		difficulty	
		P1	P2	P1	P2
BS	8.5	6.75	6.25	2.5	3.5
RS	7.5	8	4	2	2.75
BW	7.25	12.5	3.5	3.5	3
RW	4.75	11.75	3.25	3.75	2.5

Table 8. Averages showing time and difficulty when each pair appears as P1 and as P2

MAKING RECRUITMENT DECISIONS

The purpose of this experiment is to examine the ease of use and usefulness of a simple decision aid for personnel selection.

Five hypothetical candidates have made it to the shortlist for the job of Postgraduate Admissions Officer for the University. The details of the post are attached.

Each candidate has been assessed against seven criteria. The assessment is based on their application form, cv, interview, references, and performance at an assessment centre. Assessments have been scored by a panel on a scale of 1 to 5, with 5 representing excellence.

You will be presented with a table such as this

	candidate				
criterion	a	b	c	d	e
1. Written communication	2	5	4	5	4
2. Oral communication	3	1	4	2	2
3. Planning	5	3	3	4	1
4. Organising ability	5	3	4	2	1
5. Team player	4	4	2	4	2
6. Works independently	3	4	3	1	5
7. Decisiveness	4	2	4	4	2

The task is to select the preferred candidate.

An overall score for each candidate will be given by the weighted average of the scores in the table above. The candidate with the highest score will be selected.

Your input is to make statements about the relative importance of the criteria and/or the relative merits of the candidates. You may say something like

> "I think that being a team player is at least twice as important as being good at oral communication"

or

> "I think that candidate *a* is better than candidate *d*"

After each such input the score for each candidate is recalculated and displayed. You can continue with these judgements until you are satisfied with the selection recommended.

You will be asked to make two such selections, with different candidates in each case.

Figure 1. Instructions for experiment.

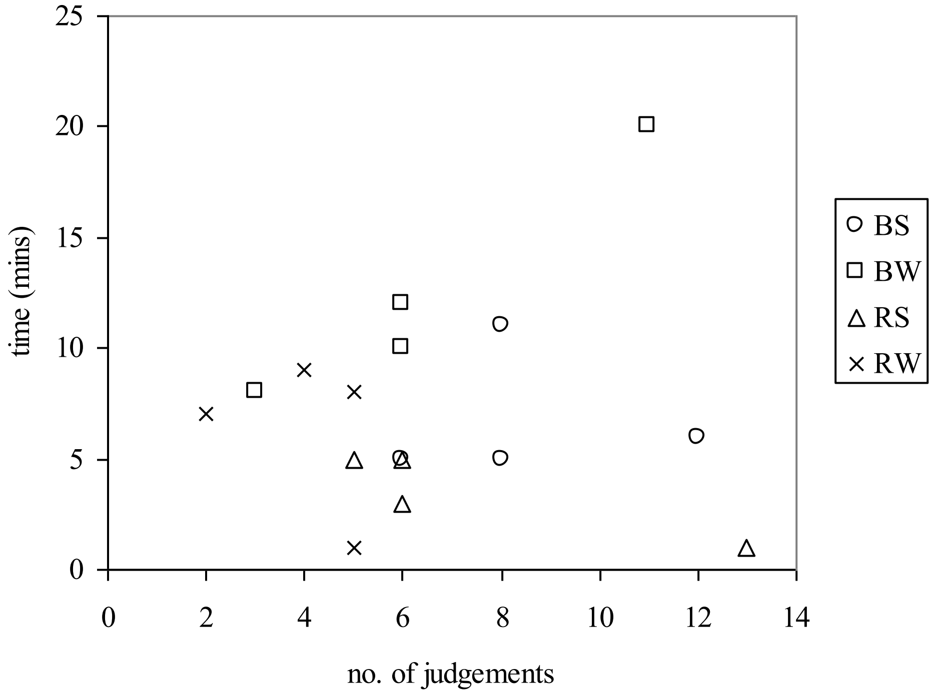

Figure 2. Time taken and no. of judgementd needed for P1 problems